



ARTICLE

<https://doi.org/10.1057/s41599-019-0299-5>

OPEN

Alien forms for alien language: investigating novel form spaces in cultural evolution

Christine Cuskley^{1,2}

ABSTRACT A cornerstone of experimental studies in language evolution has been iterated artificial language learning: studies where participants learn of artificial ‘alien’ languages, and the product of their learning is then passed onto other participants successively. Results over the last decade show that some defining features of human language can arise under these experimental conditions, which use iteration to simulate processes of cultural transmission. The form modalities used in iterated learning studies have expanded considerably in recent years, but the dynamics of how learning a *completely* novel form system interact with processes of cultural transmission are only beginning to be explored. This paper provides a brief overview of artificial language learning studies in the context of language evolution, situating them in a framework which focuses on forms used in learning and production. This issue is further explored with an iterated experiment which focuses on the role of learning completely novel forms in isolation. This experiment uses a new set of graphical symbols called Ferros, which are produced using a virtual palette. Results show that properties of this novel form space—in particular, ease of articulation—have specific effects on sequence learning and evolution. These results have implications for how forms and modalities might constrain language systems, and demonstrate how the use of truly novel alien forms might be extended to address new questions in cultural and linguistic evolution.

¹School of English Language, Literature and Linguistics and Centre for Behaviour and Evolution, Newcastle University, Newcastle upon Tyne, UK. ²Centre for Language Evolution, University of Edinburgh, Edinburgh, UK. Correspondence and requests for materials should be addressed to C.C. (email: ccuskley@gmail.com)

Introduction

Over the last two decades, cultural evolution has emerged as a framework with relevant insights into fields ranging from biology (Mesoudi, 2007) to art history (Sigaki et al. 2018), and notably, linguistics. Evolutionary linguistics was an early adopter of a cultural evolutionary framework: many concepts central to cultural evolution have been at the core of historical linguistics and language change for decades (Croft, 2006). While historical linguistics and language change have been vibrant for centuries, the study of language evolution was relatively fallow for most of the twentieth century, following the famous ‘ban’ on studying the evolution of language from the Paris linguistic society (Christiansen and Kirby, 2003); although see notable contributions such as Hockett (1960) and Hewes et al. (1973). As the field of language evolution began to solidify in the late twentieth century, the notion of an influential cultural timescale—crucially interacting with genetic and developmental timescales—became essential to empirical investigations of the evolution of language (Hurford, 1999).

This paper will focus on methods which simulate cultural evolution using an experimental approach. First, we provide an overview of the relatively short history of these methods in the context of linguistics and social psychology. This is followed by a brief discussion of how the method has been adapted to study various features of language in the broader context of how cognitive biases are amplified by cultural transmission. Finally, I present a new iterated experiment where participants produce and transmit novel forms using a two dimensional virtual palette.

Alien languages in cultural evolution. The notion of an alien encounter has often been used to frame the basic problem faced by language evolution (Hockett, 1955, NASA, 1977): assuming we want to communicate meaningfully with other cognitively advanced beings, how can we possibly do this without any pre-established system in place? In other words, how do we bootstrap a shared, mutually intelligible language from nothing? As language evolution re-emerged as a tenable problem for cognitive science in the late twentieth century, this framing has resurfaced. Depictions of alien communication in film and literature often touch upon issues central to language evolution, and the potential challenge of speaking to aliens has often helped to frame linguistic research (Little, 2018). In empirical studies of language, especially those used in language evolution, many experiments use aliens explicitly to engage participants in the task of learning novel artificial languages to communicate with a naïve interlocutor (Kirby et al. 2008; Culbertson and Schuler, 2019).

Artificial language learning (ALL) involves teaching participants miniature languages, often designed by experimenters with specific structural features in mind (e.g., plural marking). These languages usually consist of phonologically plausible words that are meaningless in the native language(s) of the participants, presented alongside meanings conveyed via images. Meanings range from simple and abstract (e.g., coloured shapes) to more concrete or complex (e.g., a child kicking a ball). Artificial language learning is often used to simulate the process of child language acquisition (Gómez and Gerken, 2000, Culbertson and Schuler, 2019), or to compare how children learn artificial languages differently from adults (Hudson Kam and Newport, 2005, Folia et al. 2010). These languages are often framed explicitly as ‘alien’, though the nonsense-words are constrained by the phonological rules of the participants’ native language (e.g., for English, *blick*, *napilu*, but not **lbick*, **ngipnu*). For adult participants in particular, the vast majority of whom are WEIRD—people from western, educated, industrialised, rich, democratic

societies (Henrich et al. 2010)—non-words are often presented in written form.

Decades before some of the earliest ALL experiments, (Bartlett, 1932) set out to study more general processes of learning and memory with distinctly cultural perspective:

The form which a rumour, or a story, or a decorative design, finally assumes within a given social group is the work of many different successive social reactions. Elements of culture, or cultural complexes, pass from person to person within a group, or from group to group, and eventually reaching a thoroughly conventionalised form, may take an established place in the general mass of culture possessed by a specific group...In this way, cultural characters which have a common origin may come to have apparently the most diverse forms. Bartlett (1932, p. 118)

To simulate this process, Bartlett used *diffusion chains* (see also, Balfour, 1893). The first step of this method involves a standard recall task, wherein the initial participant is tasked with reproducing a story or drawing after a short exposure and interval. Bartlett extended this by using the first participant’s reproduction as the target for the next participant, who underwent the same process. In this way Bartlett created ‘chains’ of participants over which drawings or stories were culturally transmitted. In short, he created an in vivo simulation of cultural transmission (Roberts, 2010). This method, and the idea of social diffusion generally, has been widely adopted in cultural evolution (Whiten and Mesoudi, 2008, Kempe and Mesoudi, 2014a, 2014b, Moussaïd et al. 2015, Caldwell et al. 2016). Kirby et al. (2008) were among the first to combine the diffusion chain method with artificial language learning specifically (see Esper (1925) and Esper (1966) for earlier efforts). Now well known as the *iterated learning model*, this method integrated artificial language learning experiments widely used in linguistics with diffusion chain experiments described by Bartlett, 1932.

Kirby et al. (2008) started with pseudo-random artificial languages¹, training participants on mappings between 2–4 syllable written forms (e.g., *napilu*), and simple shape ‘meanings’, which varied in terms of their colour, movement, and contour (e.g., a red bouncing square, a blue spinning triangle). They then instructed participants that they would be learning an ‘alien’ language. Participants were given several rounds of training where they were exposed to a subset of the mappings between the random forms and the structured meanings. In a final testing stage, they were tasked with producing forms for *all* of the meanings, even though they had only seen a subset of them. A subset of the forms they produced were then used to train a second ‘generation’ of alien language learners. This process was repeated until they had four independent ‘chains’ of ten generations each, providing miniature trajectories of language evolution.

Omitting some forms from training in each generation forms a crucial step of this process which constrains information transfer between generations, known generally as a *bottleneck*. Limiting information transfer between participants forces them to engage in generalisation during production, allowing cognitive biases to accumulate in the resulting languages. While in Kirby et al. (2008) this bottleneck was made explicit by omitting some items from training (in this case meant to be a proxy for the famous poverty of the stimulus, Zuidema, 2003), this bottleneck can take any form which results in some information loss. For example, in the original Bartlett (1932) studies, memory constraints form a bottleneck: by having short exposure to e.g., a drawing, and an interval between exposure and production (see also Tamariz and Kirby (2015)), some information is lost. More recent studies have

shown that a similar bottleneck effect can come from communication (Kirby et al. 2015) or from an expanding meaning space (Raviv et al. 2019). Using a computational approach, Spike et al. (2017) showed that some form of informational bottleneck—be it poverty of the stimulus, memory constraints (e.g., Cuskley et al. 2017), or population turnover—is generally crucial to the emergence of learned signalling systems in a population.

Kirby et al. (2008) resulted in two crucial findings. First, the miniaturised languages changed over successive ‘generations’ to become more learnable—that is, the accuracy of recall for participants in later generations was significantly higher than accuracy for early learners of the initially random ‘alien’ language. Second, the initially random languages became structured: certain parts of words referred reliably to certain features of meaning. This is a specific kind of linguistic structure known as *compositionality*. For example, in one chain the suffix *-plo* reliably recurred in words for bouncing shapes. These two results are crucially related: languages were becoming more learnable *because* they were becoming more structured, walking the fine line between systems which are easily acquired and systems which are functional for communicating diverse meanings (Carr et al. 2018).

As part of a wider drive to study cultural processes experimentally (Mesoudi, 2016, Caldwell and Millen, 2008, Kempe and Mesoudi, 2014a, 2014b), a large literature in iterated artificial language learning has emerged over the last decade. This includes direct extensions of the paradigm used by Kirby et al. Smith and Wonnacott (2010), Carr et al. (2017), Carr et al. (2018), Nölle et al. (2018), Saldana et al. (2019) and similarly innovative iterated studies which span several modalities and even species (e.g., Motamedi et al. (2018), Fehér et al. (2009), Claidière et al. (2014))—see Tamariz, (2017) for an overview. While the current discussion will focus mainly on studies related to language, note that we use the notion of forms as a general term, which could include everything from paper airplanes or spaghetti towers (Caldwell and Smith, 2012, Caldwell and Millen, 2008) to drawings (Tamariz and Kirby, 2015) and complex tools (Morgan et al. 2015). Below, we focus specifically on using an iterated approach to issues in linguistic form, which provides context to the alien form experiments to follow. In this context, the discussion below uses *forms* to refer to the labels participants learn and reproduce, *meanings* to refer to the images or other concepts labelled by the forms, and *mappings* to refer to the relationship between the two.

The role of form spaces in cultural transmission

Much of the experimental literature which considers the effect of cultural transmission on communicative systems takes one of two approaches. In experiments closely modelled on Kirby et al. (2008), participants are given meanings (e.g., shapes or concepts) and learn random (but experimenter specified) mappings between written language-like forms and those meanings. A closely related thread of studies, which emerged independently, focuses on participants in the first generation improvising forms for given meanings (e.g., drawings Garrod et al. 2010, or gesture Motamedi et al. 2018) which are passed onto later generations.

Participants in iterated artificial language learning studies tend to be fluent users of human spoken language embedded in human culture², and so even if they generate novel spoken or written forms, these are likely to be heavily constrained by their existing language systems. For example, they would inevitably generate forms which adhere to phonotactic rules in their existing language, and which may be subject to lexical neighbourhood effects (e.g., in Beckner et al. (2017), some chains evolved to use the suffix ‘-trio’ as a plural marker for groups of three items). There is

also some evidence that the emergence of structure is affected by literacy: among participants who learned a musical artificial language, structure was more likely to emerge in chains of literate musicians than other participants (Brown, 2008, Tamariz et al. 2010). The use of improvised drawings or gestures escapes this influence in some sense, but has more obvious affordances for *iconicity*: forms in the gestural or visual modality have greater potential to resemble their referents (Fay et al. 2014, Hockett, 1978). Iconicity can form an important foothold for language learning and acquisition (Imai and Kita, 2014, Dingemanse et al. 2015, Lockwood et al. 2016) and bootstrapping communication more generally (Cuskley and Kirby, 2013). Indeed, experiments show that iconic affordances often drive the initial establishment of forms, though interaction or iteration leads to their eventual conventionalisation (Theisen-White et al. 2011, Tamariz et al. 2018, Little et al. 2017a, 2017b).

These points do not make the results of Kirby et al. (2008) and related studies any less remarkable: regardless of participants’ existing biases with respect to forms, the result that structure in mappings emerges *de novo*—across written, drawn, and gestured modalities—is striking. However, by using more thoroughly ‘alien’ form spaces—which move away from written or spoken language, or the familiar iconic affordances of drawings and gestures—we can shed a different light on the emergence of symbolic forms.

The effect of an existing, acquired linguistic system on artificial language learning presents a problem for language evolution particularly. Ideally, to access the biases which constrain language *evolution*, we should aim to avoid or account for language specific biases such as phonotactic rules, lexical neighbourhood effects, and literacy. From the specific perspective of linguistic form, a broad body of research has suggested that entire classes of sound change across languages can be accounted for phonetically (Ohala, 1993, Blevins, 2004). In other words, physiological constraints of speech production and perception in humans can partially account for the distribution of sounds in languages over time and space, independent of changes which may occur in the context of language specific phonological constraints (Ohala, 1981, Ohala, 1983, Blevins, 2004). Importantly, such physiological constraints are not necessarily fixed, and may interact meaningfully with other cultural and environmental constraints (Everett et al. 2015). For example, there is evidence that, in relatively recent human history, the advent of agriculture led to developmental changes in oral physiology, which may have led to a higher frequency of dental consonants in certain languages (Blasi et al. 2019).

Overall, understanding how the physical properties of a form space affect learning and cultural transmission of forms themselves can provide insights into how these considerations effect the cultural evolution of language—particularly where, as in human languages, form spaces span spoken, written, and manual modalities. To access this experimentally, we need to move away from familiar linguistic or cultural forms—which may be subject to strong influences from previous learning—to more thoroughly ‘alien’ forms. Several lines of experimental research have focused on this potential by moving away from spoken, written, drawn, or gestured modalities. Instead, they utilise more genuinely ‘alien’ form spaces to investigate how properties of a signal space might interact with cognitive biases and task constraints to shape the structure of communicative forms.

Alien form spaces. One of the earliest efforts in this vein was Verhoef (2012), where participants copied forms generated using a slide whistle. These forms were not mapped to meanings, but were presented in isolation. Over several chains, Verhoef found

that iteration of initially random whistles resulted in increasingly *combinatorial* forms: that is, initially random forms evolved into related forms made up of re-usable sub-parts, which formed structural units. In a follow up study, Little and de Boer (2014) found that shrinking the form space by artificially shortening the slide on the whistle did not effect this overall result, showing that combinatorial properties were derived from affordances in the space aside from absolute pitch (e.g., repetition and staccato). A later study introduced mappings between whistled forms and meanings with more iconic affordances, which seemed to suppress the emergence of combinatorial structure (Verhoef et al. 2016).

A related line of research uses leap motion, a device where participants produce sounds by gesturing in particular ways within a pre-specified space (Eryilmaz and Little, 2017). Using this technique, Little et al. (2017a), (2017b) found that iconic affordances between the form and meaning spaces drive increases in form complexity, and thus, structured mappings. Similar results were obtained via repeated production and learning within a single participant (Little et al. 2017a, 2017b). More recently, Kempe et al. (2019) have used a novel tone-based language with children and adults: forms are sequences consisting of two notes mapped to simple shape meanings. Iterated transmission chains with experimenter filtering failed to generate structure, but a dyadic communicative task with more natural expressive pressures resulted in some rise in compositionality. In general, children were less successful at the task than adults, a result echoed in other iterated artificial language learning work with children (Flaherty and Kirby, 2008, Raviv and Arnon, 2018). This hints at advantages for adults in more traditional iterated ALL experiments, underscoring the importance of using more genuinely ‘alien’ forms. In other words, adults’ prior linguistic experience may provide strong scaffolding for creating structure from initially random languages. However, other task-related factors (e.g., memory, attention, training duration) may also underlie this (Kempe et al. 2019).

In each of these cases, the emergence of structured forms seems to require experimenter filtering, structure in the meaning space, or some combination of these two. In the leap motion work, Little et al. (2017a), (2017b) were specifically interested in how structure in the meaning space would be reflected in form structure, and so the structured mappings which emerged were largely driven by the relationship between the meaning space and the affordances of the form space. Likewise, earlier studies expected compositional form structure to mirror deliberate divisions in the meaning space, and have used experimental filtering to remove redundant forms (Kirby et al. 2008, Beckner et al. 2017). Other studies have used continuous meaning spaces, but find that the task of structuring this space is necessary for the emergence of structured language (Carr et al. 2018, Carr et al. 2017), alongside interesting variation in the topology of structure between chains. Although Verhoef (2012) did not have participants map whistled forms to any meanings, participants’ outputs were filtered between generations for redundant forms, creating an artificial expressivity pressure (as in Experiment 2 in Kirby et al. (2008)). A few studies have looked specifically at forms in isolation, giving an even more focused look at how forms can accumulate structure via cultural processes. In the realm of music, for example, several studies have found that initially random rhythmic or tonal sequences will acquire structure via iteration (Ravignani et al. 2016, Ravignani et al. 2018, Ravignani and Verhoef, 2018, Lumaca and Baggio, 2017), and the resulting structure is mirrored in cross-cultural musical trends (Trehub, 2015).

Cornish et al. (2013) tasked participants with learning initially random sequences of four different colours and tones³. Though

these forms were initially complex—sequences were twelve units long and contained at least one instance of each colour-tone—they were random aside from these basic constraints. Participants were tasked with simply copying a sequence with the highest fidelity possible, after very brief exposure (reminiscent of the Simon game). Over generations, the sequences not only became more learnable—despite maintaining or even extending their overall length—but also gained internal structure. This was evidenced by increased compressibility over generations, caused by the frequent re-use of easily remembered sub sequences (e.g., red-blue-red-blue).

Claidière et al. 2014 made the first attempt to perform an iterated learning experiment with artificial symbols in non-human animals. In a free-ranging captive population of baboons (*Papio papio*), Claidière et al. 2014 tested a simple iterated form copying task. Using a 4 × 4 touchscreen grid (16 cells), the baboons were tasked with copying patterns which occupied 4 random cells: they were briefly shown a pattern and had to select the same four cells from an empty grid after a short interval. By using their copied forms as input for the next ‘generation’ of baboons, the patterns were iterated over 12 generations.

Not only did copying accuracy increase over time in transmission chains, but the initially random forms converged on ‘tetrominoes’ across independent chains. In other words, the four selected cells were significantly likely to become adjacent, creating the classic shapes found in the game Tetris. In a variation on this task where innovation (selecting four *different* cells than the ones which had just been highlighted) was rewarded rather than copying, Saldana et al. (2019a) found similar results: the baboons’ performance increased over time, and there was a marked preference for producing tetrominoes. In addition, they found similar results for children, although their preferences for tetrominoes (particularly lines of 4 cells) were more extreme (see also Saldana et al. (2019b)). Kempe et al. (2015) performed a similar experiment with children and adults using a larger (10 × 10, 100 cell) grid, with no restrictions on how many cells could be filled. Both children and adults showed increased accuracy over generations and some accumulation of visual structure that differed across chains (defined in their measures as ‘identifiability’). However, adult chains maintained significantly more complexity over ‘time’ than chains of children.

Together, these results suggest two major findings. First, even under conditions where forms are not mapped to any meanings, and copying accuracy is the only performance directive for participants, forms emerge which have internal structure (Cornish et al. 2013); see also, Cornish et al. (2017). This indicates that form structure facilitates learning, even without the anchor of a structured meaning space. Saldana et al. (2019a) has slightly different implications, given that the overall complexity of forms in this case was much more constrained (i.e., they had to consist of no more or less than four cells). Given limitations on form complexity, these results suggest that constraints on ease of production are likely to form the strongest pressures on structure: the tetrominoes are likely to emerge because producing forms of adjacent cells is less effortful than selecting four non-adjacent cells (Claidière et al. 2014; although note that they did not explicitly measure this).

Both of these studies use relatively simple form spaces. The forms in Cornish et al. (2013) and Kempe et al. (2015) were not length limited, and thus could develop some internal structure that was not possible in Claidière et al. (2014). However, the form space was still relatively simple and low dimensional, consisting of one of four independent colours or the presence or absence of colour within a cell. The current paper presents an iterated form copying task similar in some respects to the ones used in earlier studies Cornish et al. (2013), Claidière et al. (2014), and Kempe

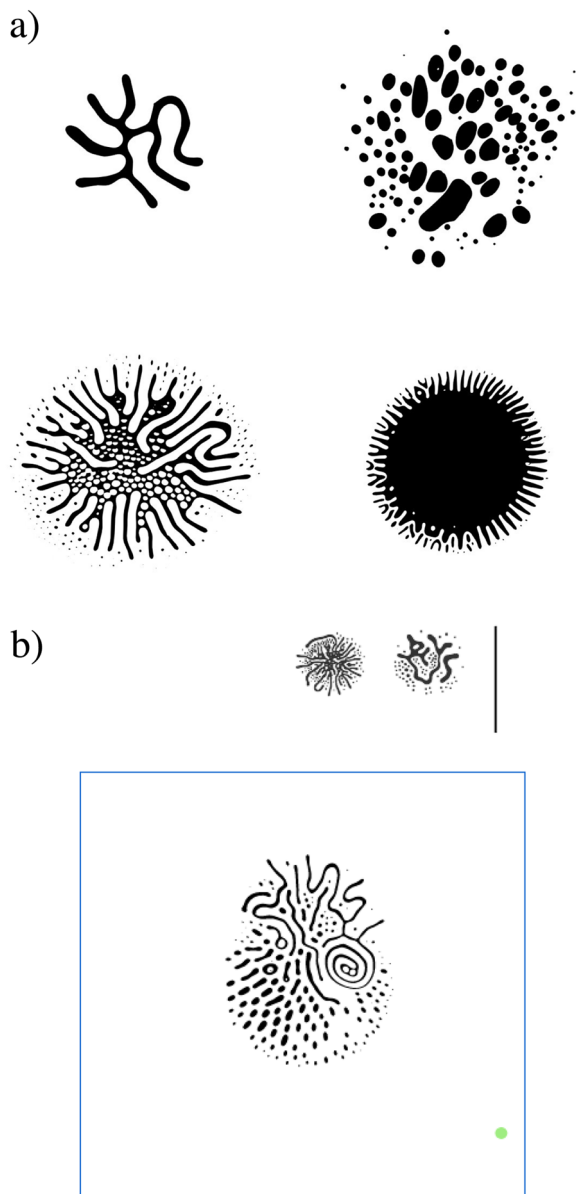


Fig. 1 a Four Ferros, corresponding within the font to the characters β , r (top row) and 5 , d (bottom row); **(b)** The virtual Ferro palette. The blue square denotes the outer edges of the palette, and the small green dot indicates the user's current position within the palette. When the user moves their mouse within the palette, the large Ferro which displays in the centre of the palette changes depending on the location of the mouse. When the user finds the Ferro symbol they wish to add to the sequence (in this case, directly above the palette), they simply click in that location

et al. (2015). However, we use a novel, complex visual form space described in detail below. Crucially, this study involves participants learning a complex form space itself *while* they use it in production, revealing ways in which emergent form systems might be constrained by the interaction between learning and use.

Ferro: truly alien forms

To further investigate the effect of form space on the evolution of structure and learnability, the study presented here uses an artificial form space which is composed of genuinely 'alien' forms, which participants will not have encountered before, and which are unlike other graphical forms they may know (e.g., letters). Furthermore, these forms are highly complex: unlike a colour-

tone pairing or a full/empty cell within a grid, each form is visually complex and comes from a larger set.

These forms are called Ferros: a set of 137 abstract graphemes visually unlike e.g., Roman orthography (see Fig. 1a). To articulate in Ferro, participants move the mouse within a square two-dimensional palette which produces a different symbol depending on the mouse's location (Fig. 1b). Distance between Ferros encodes similarity to some extent, but the forms are discrete. Unlike non-words, colours, or even random note or rhythm sequences, Ferros are signals that are *entirely* alien to participants, who have to simultaneously learn (i) to use the space where Ferros are produced and (ii) the relevant perceptual features of the Ferros themselves.

Ferro palette. Ferros were originally created as font using ferrofluid ink: ink with very small iron filings throughout it which was pressed between glass plates and exposed to different spinning magnets. This created 137 unique graphemes, each corresponding to a unicode character. Ferro was not created to be used as a traditional font, but rather, as an artistic collaboration between graphic designer Craig Ward and chemist Linden Gledhill (Ward and Gledhill, 2015). Since it was not designed to be an analogue of the Roman alphabet, it has a considerably more 'alien' quality—it has been compared to the alien glyphs from the 2016 film *Arrival*, although the font was conceived and released prior to the film. Ferro was released to backers of a Kickstarter project in late 2015 in standard .ttf and .otf formats, and thus functions like any other font: you can select it as the font in a document and type in it.

However, to use Ferro as a completely 'alien' symbol system, the current experiments aim to avoid the familiarity inherent in typing. Although participants could easily produce Ferros using the keyboard, this might make it possible for them to learn sequences of Ferros by creating mappings between Ferros and existing letters (e.g., remembering they must press the r key to produce the symbol on the top right in Fig. 1a). While this would certainly present a more difficult task than learning sequences of familiar characters, it would provide a way for participants to bootstrap their learning by mapping the Ferros directly to another culturally acquired system: letters.

Instead, we created a virtual palette used to produce Ferros. The palette is written in JavaScript using p5.js (<https://p5js.org/>)⁴. When a participant moves their mouse inside the palette, a different Ferro is produced depending on the x and y coordinates of the mouse, and the Ferro corresponding to the current mouse location (represented by the green dot in Fig. 1b) appears in large format in the centre of the palette. To choose a specific Ferro to 'write' to the field (in Fig. 1b, above the palette), the participant simply clicks when they see the desired symbol displayed in the centre.

Each Ferro symbol, like any other font, can be defined in terms of its *nodes* and *contours*. Nodes constitute control points on individual segment. For example, a straight line consists of two nodes (one at each end), while a simple curve (e.g., like the letter v) consists of at least three: one at each end, and one in the centre to create the peak of the curve which consists of two joined segments. A contour is an open or closed series of joined segments that constitute a single shape. The simple letter c shown in Fig. 2 has five nodes, and is a single contour, while a simple letter i has three nodes (each end of the line, and the dot) and two contours (one for the line, and one for the dot).

Each of the 137 Ferro symbols has a different number of nodes and contours which dictate its production location in the palette. Each Ferro was ranked in terms of its number of nodes and contours and then arranged in the two dimensional space, with

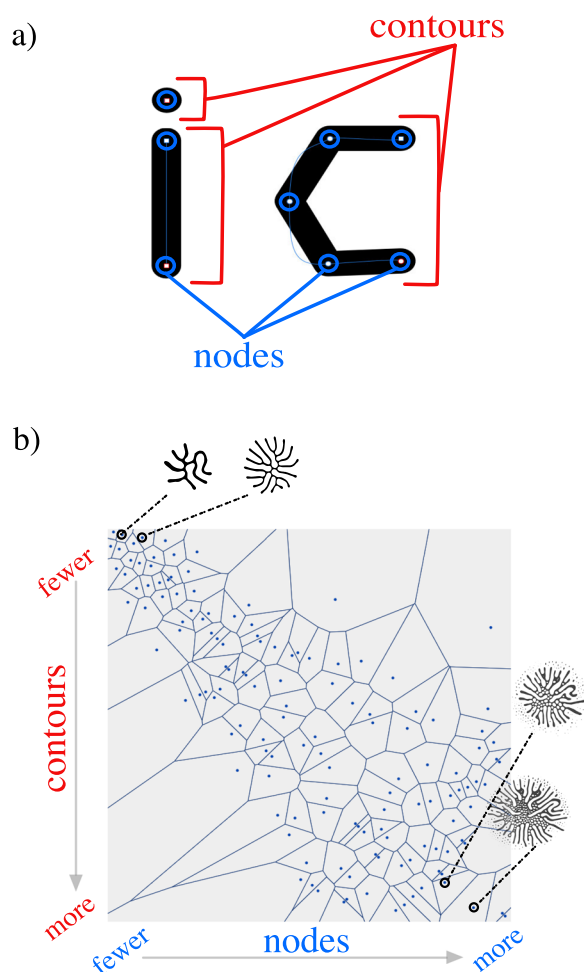


Fig. 2 **a** Nodes and contours in the context of font design. Nodes (green) can be conceptualised as control points on lines which change their shape; while this simple letter *c* has only five nodes, a *c* with a smoother shape would require more. Contours are closed shapes or segments; while the *c* has more nodes, the *i* has more contours since the dot is independent from the body. **b** The structure of the Ferro palette. Each polygon corresponds to an area within the palette that produces a different Ferro; this is determined by the fact that every point within each polygon is closer to the location of its Ferro than others (in terms of Euclidean distance). The top left shows two Ferros with relatively few nodes and contours (β : one contour and 168 nodes, and α one contour and 325 nodes), while the bottom right shows two Ferros with many nodes and many contours (γ : 230 contours and 1878 nodes, and δ : 322 contours and 2617 nodes)

the upper left corner having Ferros with the fewest nodes and contours, and the lower right containing Ferros with the most nodes and contours. While the nodes and contours for each Ferro do not increase linearly (i.e., the top ranked Ferro in terms of nodes is not necessarily top ranked in terms of contours), Ferros which have more contours do tend to have more nodes, so this space is not uniformly occupied: the upper right and lower left areas of the palette contain fewer Ferros (see polygons in Fig. 2b).

In production, the palette displays the Ferro with the shortest Euclidean distance from the location of the mouse. This is best visualised using a Voronoi tessellation, as shown in Fig. 2b. Each dot shows the central location for a Ferro, which is determined by its number of nodes and contours. While the mouse is within the polygon surrounding a given dot, that particular Ferro will be “produced” (i.e., will appear in the centre of the palette, and will

be written to the sequence field when clicked). Organising the palette in this way results in two key features. First, while each Ferro is discrete (i.e., crossing a boundary between polygons abruptly changes the symbol which is produced), physical distance and perceptual similarity are not unrelated in this space. In other words, Ferros that are closer to one another in the palette look more like each other than two distant Ferros (see proximate Ferros in the upper left and lower right in Fig. 2b).

Second, the non-uniformity of the space results in different effective surface areas of the palette for different Ferros: some Ferros occupy a sparse area in the palette (e.g., the lower left or upper right), which means that the distance the mouse can move while still producing that Ferro is relatively large. On the other hand, in more densely occupied areas of the palette (roughly, the area running from the upper left to the lower right), the mouse location has to be more specific to produce a particular Ferro. In other words, some Ferros have very large articulation spaces, while others are relatively small and specific. This means that for Ferros with large articulation spaces (i.e., large polygons shown in Fig. 2b), the mouse can be further from the “ideal” production point for a Ferro, and still produce that Ferro accurately. On the other hand, to accurately produce a Ferro with a smaller articulation space, the mouse location must be closer to the centre point.

The experiments below focus on how this second feature might effect learning and production. Participants are set with the simple task of reproducing a short sequence of three Ferros using the palette. Their productions were then used as the input for the next ‘generation’ in a standard iterated procedure. Crucially, the initial sequences copied by participants either had large articulation spaces, or small articulation spaces (Fig. 3).

In the first instance, we predict that, as in earlier iterated studies, learning will improve over time: later generations will exhibit higher accuracy (lower error) than early generations. We will look at two potential mechanisms which might underlie variation in learnability: (i) accuracy could increase over time as a result of sequences becoming simpler, and/or (ii) accuracy could increase over time as the Ferros which make up the sequences shift to those which are more robustly and accurately produced as a result of larger articulation spaces. This should be especially apparent in the small articulation space condition, where we would expect sequences to move towards Ferros with larger articulation spaces. The motivation behind these mechanisms is explained further below.

In the case of simplification (i), we should expect that a set of 12 sequences made up of only 4 unique Ferros, for example, would be easier to learn and reproduce than ones made up of 18 unique Ferros (as the initial sequences are). This expectation is well motivated by earlier iterated learning studies, where languages tend toward degeneracy without expressive or communicative pressures (Kirby et al. 2015). For example, in Experiment 1 in Kirby et al. (2008), where duplicate forms were permitted in the final testing stage and passed onto subsequent generations, this resulted in artificial languages which were highly underspecified: in one chain, there were only 5 distinct forms which applied to all 27 meanings. Here, without a pressure for mappings much less for expressivity or communication, we might expect a similar strategy for achieving learnable systems. This will be measured using entropy, which will be explained in further detail below.

In the case of production constraints (ii), we expect that learners will favour Ferros with larger articulation spaces. These Ferros are easier to find in the palette, and so may be easier to produce: they can be produced accurately by locating any point inside a relatively large polygon, rather than locating the mouse within a smaller, more exact space inside the palette. Likewise,

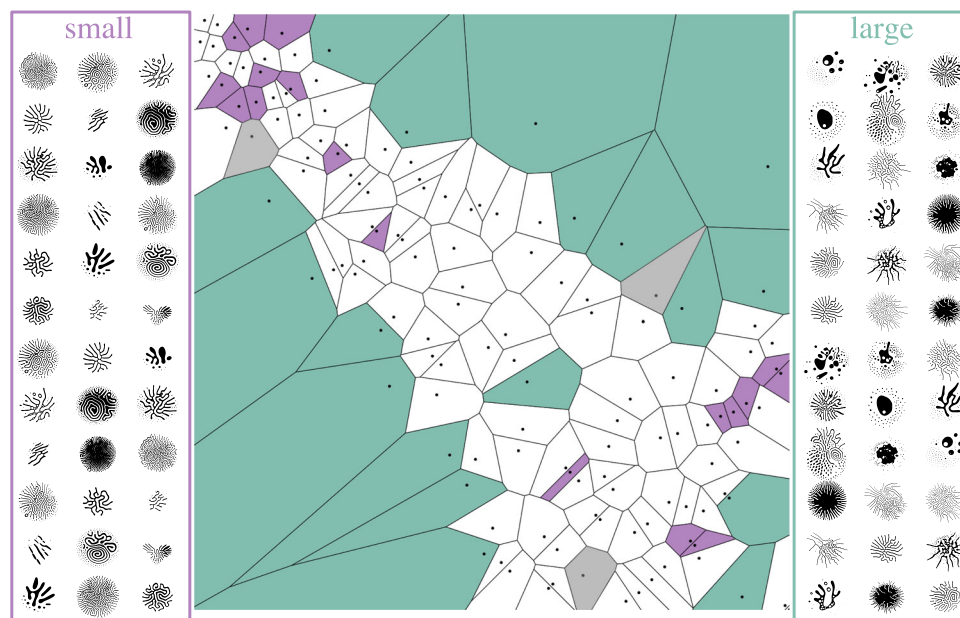


Fig. 3 Ferro sequences for the small and large articulation space conditions. Each initial condition was seeded with 12 different three Ferro sequences, which consisted of 18 unique characters. The procedure for generating the set of sequences is described in the text. Grey polygons represent the Ferros used in training sequences for all participants

some sounds in language are more “robust” than others, with single sounds actually occupying a rather large acoustic “space”: “tremendous variability exists in what we regard as the ‘same’ events in speech...the ‘same’ sound is measurably different not only when spoken by different speakers, but also when spoken by the same speaker in different phonetic environments or at different rates.” (Ohala, 1993, p. 237). Sounds which are more robustly produced and perceived, as demonstrated from biomechanical models of speech production, also tend to occur more frequently cross linguistically (Moisik and Gick, 2017). Here, we predict that a similar dynamic will make Ferros with larger articulation spaces more robustly and accurately produced across chains and generations. In other words, we predict that Ferros with larger articulation spaces have greater ease of accurate articulation, and will be more likely to be accurately learned and transmitted than Ferros with smaller articulation spaces. Accordingly, we predict that chains starting with small articulation spaces will move towards sequences with larger overall articulation spaces, while chains with larger articulation spaces are more likely to stabilize or fixate. This will be measured using surprisal, also defined in further detail below.

Learning and producing alien forms

Methods

Materials. The experiment used the Ferro palette described above, embedded in a simple HTML web page. The palette appeared to all participants as a 360 pixel by 360 pixel square. The initial sequences participants were tasked with copying were created as follows: an initial set of 18 unique Ferros was chosen for each condition based on their articulation spaces. The small articulation space Ferros occupied between 0.1 and 0.3% of the palette, while large articulation space Ferros occupied between 1 and 8% of the palette (see Fig. 3). In each condition, the 18 unique characters were put into 12 sequences with the following constraints: each sequence consisted of three unique characters, each character only appeared in two sequences, each character was in a different position in each sequence where it occurred, and was with different Ferros in each sequence where it occurred. Figure 3 shows the twelve initial sequences in each condition

Participants. Participants were recruited on Amazon Mechanical Turk. They were paid \$1.50 for the entire experiment, which generally took between 5–10 min to complete.

Procedure. The procedure of the experiment is explained in detail below, and was approved prior to the start of data collection by the Linguistics and English Language Ethics committee at the University of Edinburgh. A demo of the experiment can be viewed at <https://bit.ly/FerroSeqDemo>. Participants began the task by consenting to have their data collected for research. This was followed by a short training consisting of two phases. In the initial phase, participants were encouraged to explore and interact with the palette freely, clicking to write any Ferros they wanted to the field below the palette. Once they had selected any three characters, they were given the option to continue to the next training phase. However, they could also continue this exploration phase indefinitely, by deleting the Ferros they had produced and clicking on new ones (or simply moving through the palette to change the Ferro displayed in the centre).

In the second training phase, participants were tasked with copying training sequences, which consisted of Ferros not included in the initial sequences for either the large or small articulation space conditions (grey polygons in Fig. 3). During this training phase, participants were instructed that the goal of the task was to copy each sequence as accurately and quickly as they could—this constraint was reinforced by the fact that the target sequence slowly faded over the course of 8 seconds. This practice phase included a ‘hint circle’ which appeared after 5 seconds and slowly shrank to highlight the optimal location for the target Ferro.

Once the participant had selected three Ferros for the sequence, a ‘Submit’ button became active. Participants could delete and replace one or more Ferros at any time before pressing ‘Submit’ by pressing the backspace key. The ‘Submit’ button would be inactive when the sequence was less than three Ferros, and the palette would not write additional Ferros if the sequence was already three Ferros long, so the length of sequences was constrained. After the participant clicked ‘Submit’, the interface displayed feedback showing where the participant had made their

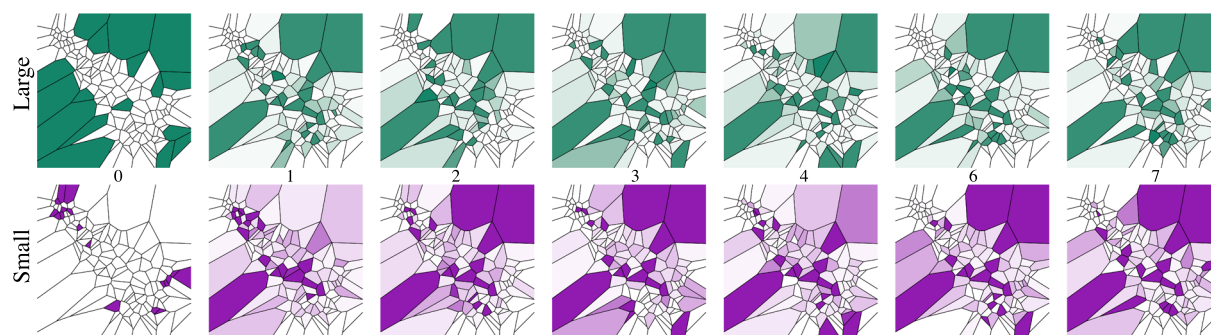


Fig. 4 Heatmaps showing how frequently each Ferro was chosen in the large (top, green) and small (bottom, purple) conditions over time. The far left panel represents the original sequences (Generation 0), with each generation progressing to the right right. Darker polygons indicate a higher frequency of a particular Ferro across chains in the relevant generation, while white polygons indicate that a Ferro was not chosen at all

selection in blue, and the target Ferro locations in orange. If the participant selected the right Ferro, the target was still displayed separately, but both the target and selection appeared in green (it is possible to choose the right Ferro without hitting the exact centre point of a Ferro's polygon). The feedback accuracy was calculated as 1 minus the normalised Euclidean distance between the target (i.e., the centre point of each polygon in Fig. 3), and the exact location of the participants' click. Participants could then click on to the next trial.

After four practice trials, participants moved to the main task, where they were informed that there would be one key difference between the training and the main task: the hint circle would not appear during the main task. For participants, target trials were otherwise identical to the practice trials. During the target trials, each chosen Ferro was recorded along with the exact coordinates of the participant's click (if participants deleted a Ferro and replaced it, the deleted click and associated Ferro were not recorded).

Participants completed all 12 sequences in a given condition in a random order, and their progress was indicated by a progress bar at the bottom of the screen. The initial sequences for each condition are presented in Fig. 3. The output for each of these sequences was then passed onto the next participant to form traditional iterated chains. To accomplish this, each generation was put on Mechanical Turk as a separate "batch", requiring a unique data point for each chain in each generation.

Analysis and results. There were 9 chains in the small condition, and 10 chains in the large condition. Each chain had 6 generations, with the exception of one chain in the large condition which had 5⁵, making for a total of 104 unique participants. The full data set is available at <https://github.com/CCuskey/AlienForms>, along with R code which details the models reported below. Figure 4 shows heatmaps of the characters chosen over time in each condition.

For the analyses, we examine three measures. First, sequence copying error over generations indicates whether the learnability of sequences increases over time as in earlier iterated studies. Accuracy is generally a preferred measure, but this requires normalised error. Although normalised Euclidean distance was used to provide feedback to participants, here we use un-normalised distances in the space to quantify error. Thus, a decrease in error over time indicates increased copying fidelity over generations.

We use two information theoretic measures to look at the potential mechanisms behind any reduction in error over generations. To assess whether the overall complexity of the Ferro sets changes (where simpler inventories of Ferros should be easier to learn), we measure the entropy of the entire set of sequences for each generation, based on the probabilities of each unique Ferro within the set. Finally, to see whether learnability

reduces as a result of ease of articulation, we measure the surprisal of the set of sequences produced in each generation. The notion of surprisal is taken from information theory (Shannon, 1948), in this case derived from the normalised area of each Ferro's polygon in the palette (formalised in Eq. (2)). Put differently, if a random point were chosen in the palette, it would be more surprising if this point fell within a small polygon than within a large one. In other words, Ferros with small articulation spaces have higher surprisal, while those with large articulation spaces have low surprisal⁶.

For each of these measures, we tested whether generation and initial condition affected the outcomes using linear mixed effects models. Each model evaluated whether the relevant measure changed meaningfully over generations, and if there was any meaningful difference between conditions. Using the method outlined in detail in Winter and Weiling (2016), each dependent measure was tested as a separate outcome variable, with generation and condition as fixed effect predictors (including the interaction between these). Condition was deviation coded, and generation was rescaled to zero. Random effects were included for trial number and participant, and random uncorrelated slopes and intercepts were included for generation and chain. Each model was fit using the `lme4` package (Bates et al. 2014), and used `step()` function in the `lmerTest()` package (Kuznetsova et al. 2018) to optimise model fit and simplicity. This model was then tested against the null model using `anova()` comparison, and significance values were obtained using the `lmerTest` package (Kuznetsova et al. 2018) with the default Satterthwaite's method. Finally, Marginal (fixed effect) and conditional (combined fixed and random effects) R^2 values were estimated using the `sem.model.fits()` function in the `piecewiseSEM` package (Lefcheck, 2016). Full code R code for the analyses is available at <https://github.com/CCuskey/AlienForms>.

Copying error. Figure 5 shows the mean copying error over time for each generation in each chain. Copying error was measured in terms of (non-normalised) Euclidean distance in the palette space between the point where the participant clicked, and the nearest edge of the polygon for the target Ferro⁷.

The model for copying error is summarised in Table 1. Using the `step()` function in the `lmerTest()` package (Kuznetsova et al. 2018), the random effects of chain and trial were removed. This model was tested with `anova()` comparison, and found to be significantly better than a null model which included only random effects ($\chi^2 = 26.93$, $df = 2$, $p < 0.001$).

The results of the model show no overall effect of generation, indicating that transmission did not necessarily result in decreased error. There was, however, a significant overall effect of condition:

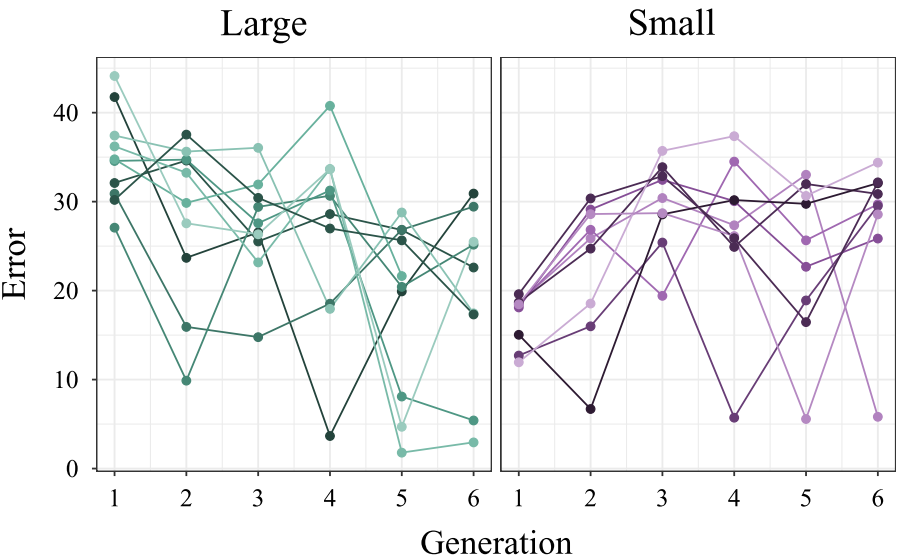


Fig. 5 Mean copying error over time for each chain in the small versus large articulation space conditions

Table 1 Summary of the model analysing error over time and between conditions				
Model: Error Generation*Condition+(0+Generation Chain) + (1 Participant)				
	Estimate	SE	t-value	p-value
(Intercept)	27.13	1.36	19.98	<0.001***
Generation	−0.76	0.46	−1.64	<0.11
Condition (large)	6.32	1.36	4.66	<0.001***
Generation × Condition (large)	−2.30	0.46	−4.98	<0.001**
Marginal R ² : 0.081, Conditional R ² : 0.34				

error was higher in the large articulation space condition. Finally, there was an interaction between generation and condition: error decreased more over generations in the large articulation space condition than in the small articulation space condition.

Sequence set entropy. The entropy was calculated across all the Ferros produced by each participant as in Eq. (1), where i is the probability of each Ferro occurring in the set produced by the participant:

$$H(X) = -\sum p(x_i) \log \frac{1}{p(x_i)}$$

(1)

Here, Entropy is expressed in *bits*: the fewer bits (i.e., the lower the entropy), the less random the Ferro set is. In this case, the lower limit would be producing the same Ferro 36 times (i.e., 12 sequences of three Ferros each, where every Ferro was the same). The upper limit would be producing 36 different Ferros, which would result in an entropy of roughly 5.17 bits. The sequences in both conditions started with the same entropy of 4.17, since they consisted of 18 Ferros, each with equal probability. Figure 6 shows the entropy over generation for each chain in each condition; the entropy of Generation 0 represents the initial sequences, while each subsequent generation represents sequences produced by a participant.

The Entropy model had slight differences from the error model: the intercept for Entropy was not random as this was fixed by the experimenter, and participant was not included as a random effect (since for this measure, each participant had only one datapoint: entropy across the entire set of Ferros produced in

the task). In the case of Entropy, the `step()` function in the `lmerTest()` package (Kuznetsova et al. 2018) dropped both of the fixed effects (Generation and Condition) and their interaction, indicating that none of these were a significant predictor of Entropy. Accordingly, the model was not a significantly better fit than an equivalent a null model, which included only random effects ($\chi^2 = 4.06$, $df = 4$, $p = 0.40$).

Ferro surprisal. The previous measure quantified predictability across the entire set of Ferros by taking a an average of the log probabilities of each Ferro produced by a participant (see Eq. (1)). However, we can also look at the log probability of a particular outcome (i.e., a particular Ferro), also known as surprisal or information content.

Given the relatively small set of Ferros produced by each participant, measuring the probability of producing a given Ferro in each generation would be minimally informative (hence the entropy approach taken above). However, we can take a different approach to probability by using the area occupied in the palette space by each Ferro: given the choice of a random point in the palette, the probability of choosing a given Ferro, x , can be defined as $p(x)$: its normalised area within the palette (i.e., the area of the polygon with x as its centroid).

$$h(x) = \log_2 \frac{1}{p(x)}$$

(2)

Surprisal is calculated using inverse probabilities (see Eq. (2)), meaning Ferros with small articulation spaces have high surprisal, while Ferros with large articulation spaces have low surprisal (i.e., it would be less surprising to choose a Ferro with a large articulation space at random). This measure is a useful test of drift in this space: if each participant randomly clicked for every sequence, we should expect the surprisal of the produced Ferros to drop over time (i.e., for sequences to tend towards being composed of Ferros with larger articulation spaces).

The sequences in both conditions started with unequal surprisal by design: surprisal was higher in the small articulation space condition, and lower in the large articulation space condition. Figure 7 shows the surprisal over generations for each chain in each condition; the surprisal of Generation 0 represents the initial sequences, while each subsequent generation represents sequences produced by a participant. The red dashed line shows the average surprisal of the entire space (equivalent to the entropy

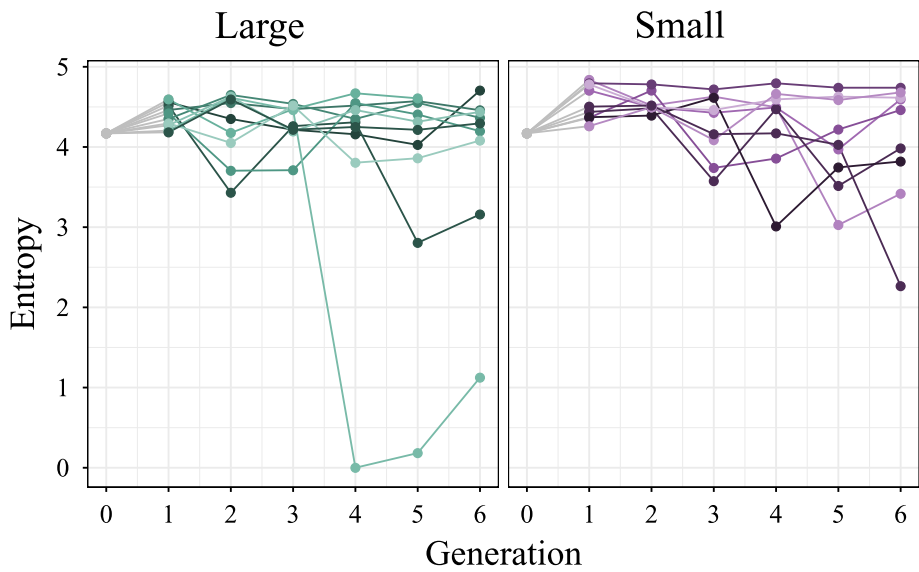


Fig. 6 Mean entropy over time for each chain in the small versus large articulation space conditions

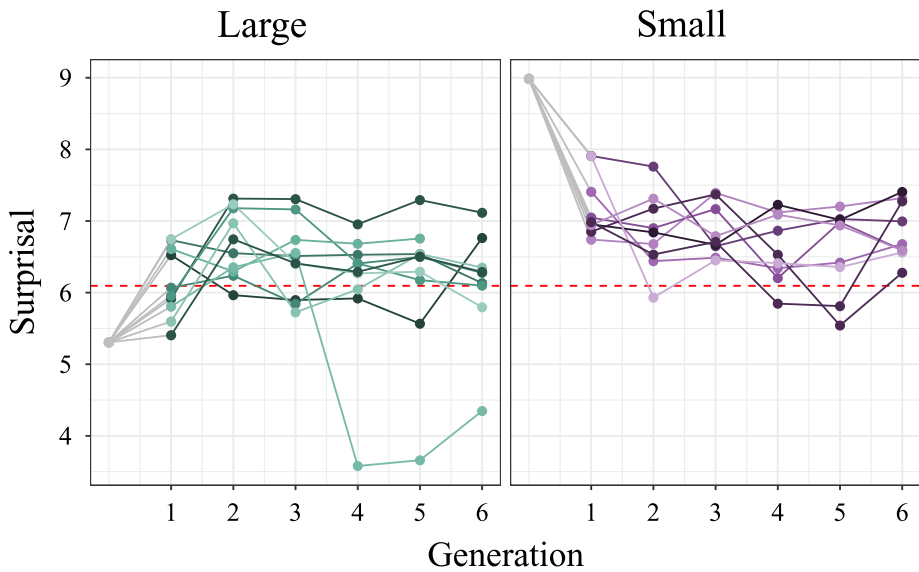


Fig. 7 Mean surprisal over time for each chain in the small versus large articulation space conditions. Red line represents the entropy of the entire space, i.e., where chains should drift if clicks were completely random

Table 2 Summary of the model analysing surprisal over time				
Model: Surprisal Condition+(1 Participant) + (0+Generation Chain)				
	Estimate	SE	t-value	p-value
(Intercept)	6.65	0.07	39.73	<0.001***
Condition (large)	−0.29	0.7	−4.235	<<0.001**

of the space); this represents where we would expect chains to converge if participants were clicking randomly.

The model for surprisal is summarised in Table 2. Using the `step()` function in the `lmerTest()` package (Kuznetsova et al., 2018), the random effects of chain and trial were removed, and the fixed effect of generation was dropped. The resulting model was tested with `anova()` comparison, and found to be significantly better than a null model which included only random effects ($\chi^2 = 5.61$, $df = 1$, $p = 0.019$).

Discussion

Several key findings emerge from the results above. The strongest findings come from the error measure, some form of which is common to almost all iterated artificial language learning studies (often in the form of accuracy). Error did not vary systematically by generation, but did vary in terms of condition, and as the result of an interaction between the two. Error was generally higher in the large condition. However, the interaction showed a decrease in error over time in the large condition (later generations had lower error). Second, the Entropy of the set of Ferros produced by each participant did not change significantly over time, nor did it differ significantly between conditions. Finally, the area-based surprisal of Ferros produced was slightly lower in the large condition, indicating that the Ferros produced in this condition tended to have larger articulation spaces. Below, we provide a brief discussion of the mechanisms which may underlie these results.

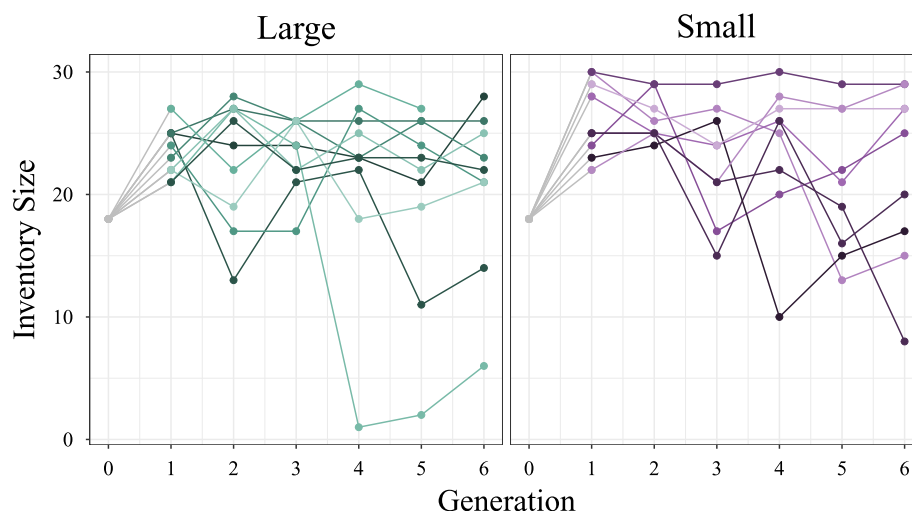


Fig. 8 Mean inventory size (number of unique Ferros) over time for each chain in the small versus large articulation space conditions

Copying error. The interaction between condition and generation in terms of error showed that copying accuracy did improve over time in the large space condition, but accuracy dropped over time in the small space condition. While higher initial error was expected in the small space condition since finding these Ferros required more precision, the rise was less expected. The main effect of condition in terms of copying error is slightly puzzling: we expected error to be lower in the large condition (because the Ferros in this condition were easier to find in the space), and generally expected error to decrease over time in both conditions.

The distribution of the large and small articulation spaces may be responsible for the higher error in the large condition in particular. Ferros with large articulation spaces tend to be adjacent to other Ferros with large articulation spaces, while small articulation space Ferros tend to be close to other small articulation space Ferros. This means that, in the large articulation space condition, participants could be further away from the edge of the correct Ferro's polygon while still producing an 'adjacent' Ferro. In contrast, adjacent Ferros in the densely occupied part of the palette are closer together in terms of Euclidean distance, which formed the core of the error measure.

In contrast, a similar distance in the more densely occupied parts of the space (where the Ferros in the small articulation space condition tended to cluster) results in more perceptual distance between Ferros. As such, the rise in error in the small space condition might reflect perceptual similarities between tightly adjacent Ferros: in the more densely occupied areas of the palette, a 'wrong' Ferro might look more like the right one (indeed, potentially almost identical, see Fig. 2b) than in the sparsely occupied areas.

Without more detailed data on how differences between Ferros are perceived, this is speculative. However, it is worth noting that perceptual boundaries—and how these interact with culturally acquired linguistic categories—are not fully understood in many domains (e.g., smell, Majid et al. 2018). This is even true to some extent of linguistic form categories: for example, some evidence shows that highly discretised vowel categories emerge early in development (e.g., Werker and Tees (2005)), while other studies show that fuzzier gradient categories form a core part of adult speech perception (e.g., McMurray et al. 2002; McMurray et al. 2018). Artificial language learning using completely novel form spaces is a fruitful avenue for further study in how shared categories simultaneously form *while* individual learners negotiate how to use those categories. Ideally, this would involve taking detailed psychophysical measurements regarding the perceived

similarity and confusability of Ferros or other novel symbols. While sufficient similarity and/or confusability ratings for the entire set of Ferros would be a substantial undertaking (a single set of pairwise comparisons between all 137 symbols would involve just over 9000 trials), future studies could use a smaller subset of the symbols to explore this issue.

Entropy. The initial sequences had relatively high entropy, consisting of 18 Ferros each with equal probability. The model for entropy showed no significant difference in this measure either over time or between conditions. However, in the context of earlier iterated studies, this in and of itself is notable. Kirby et al. (2008)'s experiments involved participants generating forms for meanings that they had never seen in training. In their initial experiment, this amounted to the expected increase in accuracy over time, but came at the cost of forms with very low entropy: participants tended to create degenerate systems where the same word was used for every meaning. Kirby et al. (2015) attribute this result to the notion that where learnability is the only pressure, this leads to a general collapse in forms: for example, a language where there is only a single form is 'ultimate' in terms of learnability. In their second experiment, Kirby et al. (2008) introduced an artificial 'expressivity' pressure by omitting duplicate forms from the training of the next generation, while later studies introduced this pressure by making the task communicative. Kirby et al. (2015) conclude that the *combination* of expressivity and learnability pressures leads to compositional structure.

In the current experiment, copying—much like learnability—was the only pressure: participants were only tasked with accurately reproducing the forms they were presented with. However, here, sets of sequences generally did not slide into degeneracy very often—at the final generation, only two chains had less than 10 unique Ferros in their set, and all but 5 chains (2 in the large space condition and 3 in the small space condition) had 20 or more unique Ferros, meaning 75% of chains actually introduced *more* items into the inventory (each language started with 18; see Fig. 8). Likewise, the vast majority of individual 3-Ferro sequences contained three unique characters (95% contained 3 unique characters, 3% contained 2, and only 2% of sequences produced consisted of a single repeated character). This suggests that although nothing in the task prevented the repetition of Ferros within a sequence, participants preferred sequences with distinct Ferros.

A similar bias to maintain surface complexity was found in an earlier form copying experiment, albeit with very low form space complexity: Cornish et al. (2013) started chains with random sequences of four colours, with the constraint that each initial sequence contained at least one instance of each colour. While sequences became more structured and learnable over time, they did not accomplish this by reducing inventory complexity: all of the sequences participants produced retained at least one instance of each colour, even though the maximally learnable sequence would have consisted of a single colour repeated a certain amount of times.

The current results show that this bias to maintain surface complexity extends to an entirely new form space that participants had to learn in the process of the experiment. This arguably makes learnability constraints even stronger in this experiment: participants in Cornish et al. (2013) would have had some experience with colours which facilitated sequence learning, but the current participants had no prior experience with Ferros. This is perhaps suggestive of some general bias in learners to maintain surface complexity, even in forms they have never encountered before, and even when such complexity is not functional or its function is unclear. This may be related to imitation biases (Legare and Nielsen, 2015), which have often been suggested as a key factor in cumulative cultural evolution (Whiten et al. 2009).

However, there is some hint that participants were not merely attempting to replicate the structure in their input, but also wished to reflect the complexity of the form space itself—even when their input was simplified. One participant introduced a completely ‘degenerate’ system consisting of a single Ferro in generation 4. Interestingly, this means that the next participant’s input was uniform both within and across sequences (which consisted of a single Ferro), but they *introduced* complexity nonetheless (as did the following participant). Future work could examine how entropy changes in a similar tasks where sequences are initially uniform (i.e., start from low entropy values instead of relatively high ones, as in the current experiment). The current results suggest that in this case, participants would introduce complexity. In the context of the task, this might be due to a bias for the complexity of produced forms to reflect not only input, but also the complexity observed in the form space. However, this constraint is likely to be replaced or at least relaxed when mappings to meanings enter the mix (as in iterated ALL studies approach). In this case, participants may shift this bias such that the complexity of forms to reflect the complexity of *meanings*, which is ultimately the driving force of emergent compositionality. Future work should focus on varying the surface complexity of initial input and form spaces themselves, examining specifically how and when this complexity is maintained, increased, or lost.

Surprisal. The surprisal measure was based on the probability of randomly producing a given Ferro based on its area within the palette: randomly choosing a Ferro which has a small articulation space would be fairly surprising (i.e., have high surprisal), while randomly producing a Ferro with a large articulation space would be less surprising (i.e., have low surprisal). We expected languages might converge to larger articulation spaces over time, generation was not a significant predictor of surprisal. Only condition significantly predicted surprisal, with the large articulation space condition having generally lower surprisal (i.e., consisting of Ferros with generally larger articulation spaces). Although the initial surprisal values were not included in the model—since these varied systematically between conditions by design—these results are likely to have been driven by the initial languages.

Overall, surprisal rose from its initial value in the large space condition, and dropped in the small articulation space condition,

with both conditions tending towards values slightly above the average surprisal in the space. We expected languages to move towards larger articulation spaces primarily as a result of ease of articulation: Ferros with larger articulation spaces (lower surprisal) would be easier to produce within the palette, and would thus be reproduced more accurately and transmitted more faithfully across generations. However, this result indicates that some form of drift might be particularly influential in this task, and this has been recently identified as a potentially influential force in language evolution (Newberry et al. 2017). But a closer look shows that participants are not *only* choosing the Ferros with the largest articulation spaces, indicating that factors other than drift related to the probability of randomly producing a given Ferro were at play. Figure 9 shows a heatmap of all Ferros collapsed across generations in each condition alongside an area-based heatmap.

These heatmaps were created by subtracting the normalised area (which is equivalent to the probability of a random click producing a given Ferro, $P(x)$ in Eq. (2)) from the production frequency across all chains and generations in the condition (i.e., the probability of a Ferro having been produced in the task itself). Ferros with the largest articulation spaces are generally under-represented in both conditions (i.e., they were produced less than we would have expected were clicks random), even though many of these were present in the initial sequences for the large space condition (Fig. 9). Across both conditions, Ferros with relatively small articulation spaces were produced more than would be expected if clicks were random, even though most of these Ferros were not present in the initial sequences. In particular, there seems to have been some bias towards Ferros in the centre of the palette in both conditions: in both the small and large conditions, the Ferro produced by placing the mouse directly in the centre was over represented. Overall, both conditions seemed to prefer the more densely populated areas of the palette, leading to an under-representation of Ferros with large spaces and an over-representation of Ferros with small spaces.

The source of these apparent preferences requires further study. It may be that the Ferro in the centre of the palette emerged in both chains because participants often began their search in the centre of the palette. However, the centre of the palette also requires some additional effort to reach: after advancing the trial outside the palette, participants would have had to re-enter the palette and cross into the centre, producing other Ferros in the process. Moreover, the corners of the palette seemed to be avoided, even though these are easy to locate and would have resulted in more reliable and accurate production than Ferros elsewhere in the palette. In all, this highlights how unexpected preferences for in alien form spaces might lead to particular trends in emerging systems that cannot be observed in tasks which use more familiar domains.

Conclusions. This study provided the first test of iterated learning using a truly ‘alien’ form space: participants had no knowledge of the signals produced by the palette, and had to learn how the palette worked—including the topology of the space—‘on the fly’ as they engaged in a sequence copying task. Results showed that error decreased over in the large articulation space in particular. Surprisingly, the entropy of Ferros in a sequence set did not change over time, despite previous iterated studies showing a tendency towards degeneracy without pressures for expressivity in the task (either artificially introduced by experimenters, or built-in as a result of communication). Finally, there was a difference in the Ferros that tended to be produced across conditions, with chains in the large condition generally preferring Ferros with larger articulation spaces. Both conditions systematically under-produced many of the Ferros with the largest articulation spaces, and showed a

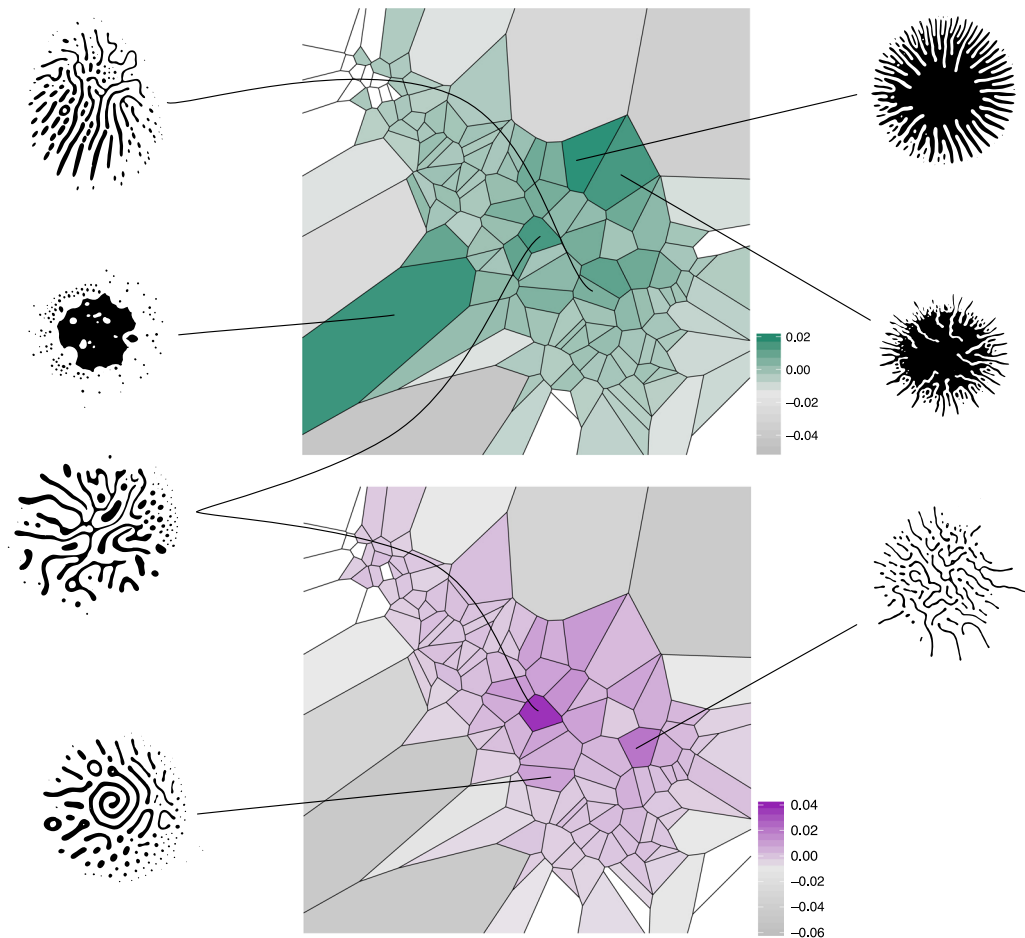


Fig. 9 Heatmaps representing the extent to which Ferros chosen in each condition were selected more than would have been expected by chance, based on subtracting the normalised area (i.e., $P(x)$ in Eq. (2)) from the production frequency across all chains and generations in the condition. More saturated purple (small condition) or green (large condition) indicates a Ferro which was selected more than would have been expected by chance. Grey polygons indicate Ferros that were underselected (i.e., based on their area, we would have expected them to be chosen more than they were). White polygons indicate Ferros that were not chosen at all

broad preference for the more densely occupied parts of the space, particularly the palette's centre.

Further work is needed to understand Ferros specifically: the perceptual features of Ferros may have played a role in important ways. In the case of error, a better understanding of the relationship between Euclidean distance in the palette space and perceptual distance between Ferros is needed. In terms of which Ferros tend to be under-produced or over-produced relative to their expected frequency based on their production area in the palette, other constraints may underlie these preferences. Participants may prefer Ferros which are more easily distinguished (e.g., have an overall darker appearance), or have readily identifiable features (e.g., a spiral).

How participants categorise Ferros in the space, and the strategies participants use to remember Ferros, both warrant further exploration. A fruitful avenue for investigating this may be to use Ferros as *meanings* in an ALL context rather than forms. The current paper focused on issues in learning, producing, and transmitting forms. Other work in iterated artificial language learning has investigated how participants divide up meaning spaces (Silvey et al. 2015, Carr et al. 2017) using more traditional written forms. A study where Ferros are used as meanings paired with more traditional written forms could shed significant light on how participants partition this novel space.

While the current results are specific to Ferros, they highlights the importance of exploring iterated learning and cultural evolution of truly novel, 'alien' forms and systems. The act of having to learn a form space while simultaneously learning to *use* it could have crucial effects on the structure of communicative forms, especially as these emerge as part of shared cultural systems. Future studies should aim to explore this using different kinds of alien form spaces, spanning modalities (e.g., visual vs. auditory), architectures (e.g., continuous vs. discrete), and contexts (e.g., dyadic and group communication). A better understanding of how we structure completely alien form spaces as we learn them can shed new light on how shared cultural and communicative systems emerge.

Received: 1 March 2019 Accepted: 11 July 2019

Published online: 06 August 2019

Notes

- 1 These languages were *structurally* random, but still phonologically possible English words, which used a finite set of syllables. In other words, they were not *completely* random strings.
- 2 There are some notable exceptions, in particular (Claidière et al. 2014), which is described in more detailed below.

- 3 Note that these were not *mappings* between colours and tones—the same colours always occurred with the same tones, and they were never presented independently—colour and tone were redundant cues.
- 4 Full code for the palette is available at <https://github.com/CCuskley/AlienForms>, although note that this version uses a freely available ‘alien’ font (albeit with the node and contour values determined by the Ferros presented here). The Ferrofluid font itself is licensed to Ward and Gledhill (2015). The artists have granted permission for the author to share the font with other researchers upon request. If you wish to use it for research, simply contact the author.
- 5 Although 10 chains of 6 generations were planned for each condition, one in the small condition stalled at generation 1, while one in the large condition stalled at generation 5 (the former is excluded entirely, while the latter provided enough data to examine some change over ‘time’). This encoding error meant that the output of the generation could be decoded by the experimenter, but not read by the browser of a subsequent participant, and so caused a termination point for the chain.
- 6 Note that, unlike the error measure, these measures are agnostic regarding the distances between forms. This means that while a change in entropy or surprisal between generations means that a participant generated a different sequence than their input, a static entropy or surprisal value might indicate the same thing: in other words, it is possible to produce a different sequence with very similar entropy or surprisal values.
- 7 Note that this is slightly different from the feedback participants received during the task. Since feedback was provided in real-time within the browser, the less computationally intensive value of normalised Euclidean distance was used (in this case, the distance was normalised by the furthest possible point from the target that was still within the palette). The calculation of distance from the click to the nearest polygon edge is a more difficult problem (involving finding the point at which each line between the click and each edge intersects, calculating the distance between this intersection and the click, and then choosing the minimum). This was done in post-processing using the Shapely package in Python (code for which is located in the Github repository: <https://github.com/CCuskley/AlienForms>). Doing this in real-time during the experiment is not impossible, but was avoided since it would have slowed the task significantly from the participant’s perspective.

References

- Balfour H (1893) *The evolution of decorative art*. Macmillan, New York
- Bartlett F (1932) *Remembering: a study in experimental and social psychology*. Cambridge University Press, Cambridge, UK
- Bates D, Maechler M, Bolker B, Walker S (2014) lme4: Linear mixed-effects models using Eigen and S4. <http://CRAN.R-project.org/package=lme4>
- Beckner C, Pierrhumbert J, Hay J (2017) The emergence of linguistic structure in an online iterated learning task. *J Lang Evol* 2:160–176
- Blasi DE, Moran S, Moisi SR, Widmer P, Dediu D, Bickel B (2019) Human sound systems are shaped by post-neolithic changes in bite configuration. *Science* 363:eaav3218
- Blevins J (2004) *Evolutionary phonology: the emergence of sound patterns*. Cambridge University Press, Cambridge, UK
- Brown E (2008) *Literacy, linguistics and compositionality: investigating the effects of cultural systems on learning and language*. Master’s thesis, University of Edinburgh
- Caldwell C, Millen AE (2008) Experimental models for testing hypotheses about cumulative cultural evolution. *Evol Hum Behav* 29:165–171
- Caldwell C, Smith K (2012) Cultural evolution and perpetuation of arbitrary communicative conventions in experimental microsocieties. *PLoS ONE* 7:e43807
- Caldwell CA, Atkinson M, Renner E (2016) Experimental approaches to studying cumulative cultural evolution. *Curr Dir Psychol Sci* 25:191–195
- Carr J, Smith K, Cornish H, Kirby S (2017) The cultural evolution of structured languages in an open ended, continuous world. *Cognitive Science* 41:892–923
- Carr J, Smith K, Culbertson J, Kirby S (2018) Simplicity and informativeness in semantic category systems. *PsyArXiv*
- Christiansen MH, Kirby S (2003) Language evolution: consensus and controversies. *Trends Cogn Sci* 7:300–307
- Claidière N, Smith K, Kirby S, Fagot J (2014) Cultural evolution of systematically structured behaviour in a non-human primate. *Proc R Soc B* 281:20141541
- Cornish H, Dale R, Kirby S, Christiansen MH (2017) Sequence memory constraints give rise to language-like structure through iterated learning. *PLOS ONE* 12:1–18. <https://doi.org/10.1371/journal.pone.0168532>
- Cornish H, Smith K, Kirby S (2013) Systems from sequences: an iterated learning account of the emergence of systematic structure in a non-linguistic task. In: Knauff M, Pauen M, Sebanz N, Wachsmuth I (eds) *Proceedings of the 35th Annual Conference of the Cognitive Science Society*. Cognitive Science Society, Austin, p 340–345
- Croft W (2006) The relevance of an evolutionary model to historical linguistics. *Amst Stud theory Hist Linguist Sci, Ser 4* 279:91
- Culbertson J, Schuler K (2019) Artificial language learning in children. *Annu Rev Linguist* 5:353–373
- Cuskley C, Castellano C, Colaiori F, Loreto V, Pugliese M, Tria F (2017) The regularity game: Investigating linguistic rule dynamics in a population of interacting agents. *Cognition* 159:25–32
- Cuskley C, Kirby S (2013) Synesthesia, cross-modality, and language evolution. In: Simner J, Hubbard E (eds) *Oxford Handbook of Synesthesia*. Oxford University Press, Oxford, UK p 869–907
- Dingemanse M, Blasi DE, Luyyan G, Christiansen MH, Monaghan P (2015) Arbitrariness, iconicity, and systematicity in language. *Trends Cogn Sci* 19:603–615
- Eryilmaz K, Little H (2017) Using leap motion to investigate the emergence of structure in speech and language. *Behav Res Methods* 49:1748–1768
- Esper E (1925) A technique for the experimental investigation of associative interference in artificial linguistic material. *Linguistic Society of America. Language Monographs*, 1:1–47
- Esper E (1966) Social transmission of an artificial language. *Language* 42:575–580
- Everett C, Blasi DE, Roberts SG (2015) Climate, vocal folds, and tonal languages: connecting the physiological and geographic dots. *Proc Natl Acad Sci* 112:1322–1327
- Fay N, Lister C, Ellison T, Goldin-Meadow S (2014) Creating a communication system from scratch: gesture beats vocalization hands down. *Front Psychology* 5:354
- Fehér O, Wang H, Saar S, Mitra PP, Tchernichovski O (2009) De novo establishment of wild-type song culture in the zebra finch. *Nature* 459:564
- Flaherty M, Kirby S (2008) Iterated language learning in children. In: Smith A, Smith K, Ferrer i Cancho R (eds) *Proceedings of the 7th International Conference on Language Evolution, EvoLang7*. World Scientific, Singapore, p 425–427
- Folia V, Uddén J, De Vries M, Forkstam C, Petersson KM (2010) Artificial language learning in adults and children. *Lang Learn* 60:188–220
- Garrod S, Fay N, Rogers S, Walker B, Swoboda N (2010) Can iterated learning explain the emergence of graphical symbols? *Interact Stud* 11:33–50
- Gómez RL, Gerken L (2000) Infant artificial language learning and language acquisition. *Trends Cogn Sci* 4:178–186
- Henrich J, Heine SJ, Norenzayan A (2010) Most people are not weird. *Nature* 466:29
- Hewes GW, Andrew RJ, Carini L, Choe H, Gardner RA, Kortlandt A, Krantz GS, McBride G, Nottebohm F, Pfeiffer J, Rumbaugh DG, Steklis HD, Ralieg MJ, Stopa R, Suzuki A, Washburn SL, Wescott RW (1973) Primate communication and the gestural origin of language. *Curr Anthropol* 14:5–24. <https://doi.org/10.1086/201401>
- Hockett C (1955) How to learn martian. *Astounding Science Fiction* 55:76–84
- Hockett C (1960) The origin of speech. *Sci Am* 203:88–97
- Hockett C (1978) In search of jove’s brow. *Am Speech* 53:243–313
- Hudson Kam CL, Newport EL (2005) Regularizing unpredictable variation: the roles of adult and child learners in language formation and change. *Lang Learn Dev* 1:151–195
- Hurford J (1999) The evolution of language and languages. In: Dunbar R, Knight C, Power C (eds) *The evolution of culture*. Edinburgh University Press, Edinburgh, UK, p 173–193
- Imai M, Kita S (2014) The sound symbolism bootstrapping hypothesis for language acquisition and language evolution. *Philos Trans R Soc B* 369:20130298
- Kempe M, Mesoudi A (2014a) Experimental and theoretical models of human cultural evolution. *Wiley Interdiscip Rev* 5:317–326
- Kempe M, Mesoudi A (2014b) An experimental demonstration of the effect of group size on cultural accumulation. *Evol Hum Behav* 35:285–290
- Kempe V, Gauvrit N, Forsyth D (2015) Structure emerges faster during cultural transmission in children than in adults. *Cognition* 136:247–254
- Kempe V, Gauvrit N, Gibson A, Jamieson M (2019) Adults are more efficient in creating and transmitting novel signalling systems than children. *J Lang Evol* 4:44–70. <https://doi.org/10.1093/jole/lzy012>
- Kirby S, Cornish H, Smith K (2008) Cumulative cultural evolution in the laboratory: An experimental approach to the origins of structure in human language. *Proc Natl Acad Sci* 105:10681–10686. <https://doi.org/10.1073/pnas.0707835105>
- Kirby S, Tamariz M, Cornish H, Smith K (2015) Compression and communication in the cultural evolution of language. *Cognition* 141:87–102
- Kuznetsova A, Brockhoff PB, Christensen RHB (2018) lmerTest: Tests in Linear Mixed Effects Models. <https://CRAN.R-project.org/package=lmerTest>
- Lefcheck JS (2016) piecewissem: Piecewise structural equation modelling in r for ecology, evolution, and systematics. *Methods Ecol Evol* 7:573–579
- Legare CH, Nielsen M (2015) Imitation and innovation: the dual engines of cultural learning. *Trends Cogn Sci* 19:688–699
- Little H (2018) Babel-on-5, Babel: The Language Magazine, 23 p 42–44
- Little H, de Boer B (2014) The effect of size of articulation space on the emergence of combinatorial structure. In: Cartmill E, Roberts S, Lyn H, Cornish H (eds) *The Evolution of Language: Proceedings of the 10th international conference (EvoLangX)*. World Scientific, Singapore, p 479–481

- Little H, Eryilmaz K, de Boer B (2017a) Conventionalisation and discrimination as competing pressures on continuous speech-like signals. *Interact Stud* 18:352–375
- Little H, Eryilmaz K, de Boer, B (2017b) Signal dimensionality and the emergence of combinatorial structure. *Cognition* 168:1–15. <https://doi.org/10.1016/j.cognition.2017.06.011>
- Lockwood G, Dingemans M, Hagoort P (2016) Sound-symbolism boosts novel word learning. *J Exp Psychol* 124:1274
- Lumaca M, Baggio G (2017) Cultural transmission and evolution of melodic structures in multi-generational signaling games. *Artif life* 23:406–423
- Majid A, Burenhult N, Stensmyr M, De Valk J, Hansson BS (2018) Olfactory language and abstraction across cultures. *Philos Trans R Soc B* 373:20170139
- McMurray B, Danelz A, Rigler H, Seedorff M (2018) Speech categorization develops slowly through adolescence. *Dev Psychol* 54:1472
- McMurray B, Tanenhaus MK, Aslin RN (2002) Gradient effects of within-category phonetic variation on lexical access. *Cognition* 86:B33–B42
- Mesoudi A (2007) Biological and cultural evolution: similar but different. *Biol Theory* 2:119–123
- Mesoudi A (2016) Cultural evolution: integrating psychology, evolution and culture. *Curr Opin Psychol* 7:17–22
- Moisik SR, Gick B (2017) The quantal larynx: the stable regions of laryngeal biomechanics and implications for speech production. *J Speech Lang Hear Res* 60:540–560
- Morgan TJ, Uomini NT, Rendell LE, Chouinard-Thuly L, Street SE, Lewis HM, Cross CP, Evans C, Kearney R, de la Torre I et al. (2015) Experimental evidence for the co-evolution of hominin tool-making teaching and language. *Nat Commun* 6:6029
- Motamedi Y, Schouwstra M, Culbertson J, Kirby S (2018) The emergence of spatial modulation in artificial sign languages. *PsyArXiv*
- Moussaïd M, Brighton H, Gaissmaier W (2015) The amplification of risk in experimental diffusion chains. *Proc Natl Acad Sci* 112:5631–5636
- NASA/JPL (1977) The Voyager Golden Record. Retrieved from: <https://voyager.jpl.nasa.gov/golden-record/>
- Newberry MG, Ahern CA, Clark R, Plotkin JB (2017) Detecting evolutionary forces in language change. *Nature* 551:223
- Nölle J, Staib M, Fusaroli R, Tylen K (2018) The emergence of systematicity: how environmental and communicative factors shape a novel communication system. *Cognition* 181:93–104
- Ohala JJ (1981) Articulatory constraints on the cognitive representation of speech. *Advances in Psychology* 7, p 111–122
- Ohala JJ (1983) The Origin of Sound Patterns in Vocal Tract Constraints. In: MacNeilage P.F. (eds) *The Production of Speech*. Springer, New York, NY, p 289–216
- Ohala JJ (1993) The phonetics of sound change. In: Charles Jones (ed.) *Historical Linguistics: Problems and Perspectives*. London, Longman, p 237–278
- Ravignani A, Delgado T, Kirby S (2016) Musical evolution in the lab exhibits rhythmic universals. *Nat Hum Behaviour* 1:0007. <https://doi.org/10.1038/s41562-016-0007>
- Ravignani A, Thompson B, Grossi T, Delgado T (2018) Evolving building blocks of rhythm: how human cognition creates music via cultural transmission. *Annals New Y Acad Sci* 1423:176–187. <https://doi.org/10.1111/nyas.13610>
- Ravignani A, Verhoef T (2018) Which melodic universals emerge from repeated signaling games? a note on lumaca and baggio (2017). *Artif life* 24:149–153
- Raviv L, Arnon I (2018) Systematicity, but not compositionality: examining the emergence of linguistic structure in children and adults using iterated learning. *Cognition* 181:160–173
- Raviv L, Meyer A, Lev-Ari S (2019) Compositional structure can emerge without generational transmission. *Cognition* 182:151–164
- Roberts G (2010) An experimental study of social selection and frequency of interaction in linguistic diversity. *Interact Stud* 11:138–159
- Saldana C, Fagot J, Kirby S, Smith K, Claidière N (2019a) High-fidelity copying is not necessarily the key to cumulative cultural evolution: a study in monkeys and children. *Proc R Soc B* 286:20190729. <https://doi.org/10.1098/rspb.2019.0729>
- Saldana C, Kirby S, Truswell R, Smith K (2019) Compositional hierarchical structure evolves through cultural transmission: an experimental study. *J Lang Evol* <https://doi.org/10.1093/jole/lzz002>
- Saldana C, Smith K, Kirby S, Fagot J, Claidière N (2019b) Cumulative cultural evolution in a non-copying task in children and Guinea baboons. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, available at: <https://mindmodeling.org/cogsci2019/papers/0186/0186.pdf>
- Shannon CE (1948) A mathematical theory of communication. *Bell Syst Tech J* 27:379–423
- Sigaki HY, Perc M, Ribeiro HV (2018) History of art paintings through the lens of entropy and complexity. *Proc Natl Acad Sci* 115:E8585–E8594
- Silvey C, Kirby S, Smith K (2015) Word meanings evolve to selectively preserve distinctions on salient dimensions. *Cogn Sci* 39:212–226
- Smith K, Wonnacott E (2010) Eliminating unpredictable variation through iterated learning. *Cognition* 116:444–449
- Spike M, Stadler K, Kirby S, Smith K (2017) Minimal requirements for the emergence of learned signaling. *Cogn Sci* 41:623–658
- Tamariz M (2017) Experimental studies on the cultural evolution of language. *Annu Rev Linguist* 3:389–407
- Tamariz M, Brown J, Murray K (2010) The role of practice and literacy in the evolution of linguistic structure. In: Smith A, Schouwstra M, de Boer B, Smith K (eds) *The Evolution Of Language: Proceedings of the 8th International Conference, World Scientific, Singapore*, p 313–320
- Tamariz M, Kirby S (2015) Culture: copying, compression, and conventionality. *Cogn Sci* 39:171–183
- Tamariz M, Roberts S, Isidro Martinez J, Santiago J (2018) The interactive origin of iconicity. *Cogn Sci* 42:334–349
- Theisen-White C, Kirby S, Oberlander J (2011) Integrating the horizontal and vertical cultural transmission of novel communication systems. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, <https://escholarship.org/content/qt40t7578c/qt40t7578c.pdf>
- Trehub SE (2015) Cross-cultural convergence of musical features. *Proc Natl Acad Sci* 112:8809–8810
- Verhoef T (2012) The origins of duality of patterning in artificial whistled languages. *Lang Cogn* 4:357–380
- Verhoef T, Walker E, Marghetis T (2016) Cognitive biases and social coordination in the emergence of temporal language. In: *Proceedings of the Annual Meeting of the Cognitive Science Society*, <https://mindmodeling.org/cogsci2016/papers/0451/paper0451.pdf>
- Ward C, Gledhill L (2015) Fe2o3 glyphs-a conceptual, ornamental type system. *Kickstarter* <https://www.kickstarter.com/projects/mrcraigward/fe2o3-glyphs-a-conceptual-ornamental-type-system>
- Werker JF, Tees RC (2005) Speech perception as a window for understanding plasticity and commitment in language systems of the brain. *Dev Psychobiol* 46:233–251
- Whiten A, McGuigan N, Marshall-Pescini S, Hopper LM (2009) Emulation, imitation, over-imitation and the scope of culture for child and chimpanzee. *Philos Trans R Soc B* 364:2417–2428
- Whiten A, Mesoudi A (2008) Establishing an experimental science of culture: animal social diffusion experiments. *Philos Trans R Soc B* 363:3477–3488
- Winter B, Weiling M (2016) How to analyze linguistic change using mixed models, growth curve analysis, and generalised additive modeling. *J Lang Evol* 1:7–18
- Zuidema W (2003) How the poverty of the stimulus solves the poverty of the stimulus. In: *Advances in neural information processing systems*, ed. Becker S, Thrun S, & Obermayer K, p 51–58. MIT Press

Acknowledgements

This work was funded by a British Academy Postdoctoral Fellowship, grant no. pf150065. Special thanks to Sean Roberts, Stefan Hartmann, Hannah Little, and Michael Pleyer for feedback in piloting the experiment.

Additional information

The online version of this article (<https://doi.org/10.1057/s41599-019-0299-5>) contains supplementary material, which is available to authorised users.

Competing interests: The author declares no competing interests.

Reprints and permission information is available online at <http://www.nature.com/reprints>

Publisher's note: Springer Nature remains neutral with regard to jurisdictional claims in published maps and institutional affiliations.



Open Access This article is licensed under a Creative Commons Attribution 4.0 International License, which permits use, sharing, adaptation, distribution and reproduction in any medium or format, as long as you give appropriate credit to the original author(s) and the source, provide a link to the Creative Commons license, and indicate if changes were made. The images or other third party material in this article are included in the article's Creative Commons license, unless indicated otherwise in a credit line to the material. If material is not included in the article's Creative Commons license and your intended use is not permitted by statutory regulation or exceeds the permitted use, you will need to obtain permission directly from the copyright holder. To view a copy of this license, visit <http://creativecommons.org/licenses/by/4.0/>.